Nonparametric estimate of the ruin probability in a pure-jump Lévy risk model

Hailiang Yang

Department of Statistics and Actuarial Science
The University of Hong Kong

Based on a paper with Zhimin Zhang

Tokyo, September 2013

Ruin theory: a few words about the history

- Lundberg, F. (1903). *Approximerad Framstallning av Sannolikhetsfunktionen*
- Cramér H. (1930). *On the Mathematical Theory of Risk*
- Cramér, H. (1955). Collective risk theory: A survey of the theory from the point of view of the theory of stochastic process
- Gerber, H. U. and Shiu, E. S. W. (1998). On the time value of ruin

Ruin theory: a few words about the literature

- Harald Cramér (1969) stated that *Filip Lundberg's works on risk theory were all written at a time when no general theory of stochastic processes existed, and when collective reinsurance methods, in the present day sense of the word, were entirely unknown to insurance companies. In both respects his ideas were far ahead of his time, and his works deserve to be generally recognized as pioneering works of fundamental importance.*

- The Lundberg inequality is one of the most important results in risk theory. Lundberg's work was not rigorous in terms of mathematical treatment. Cramér (1930, 1955) provided a rigorous mathematical proof of the Lundberg ineqality. Nowadays, popularly used methods in ruin theory are renewal theory and the martingale method. The former emerged from the work of Feller (1968, 1971) and the latter came from that of Gerber (1973, 1979).

Why insurance companies do not use the results from ruin theory

- Answer: The models in ruin theory are too simple or do not fit the real data

Classical insurance risk models

$$U(t) = u + c \cdot t - S(t),$$

where

$$S(t) = \sum_{i=1}^{N(t)} X_i,$$

$N(t)$ Poisson process, X_i i.i.d sequence.

Study of ruin probability

- Closed formula
- Upper bound
- Asymptotic results

Other ruin quantities

- Surplus before and after ruin
- Gerber-Shiu function
- ...

Highlights of this paper

- Propose a nonparametric estimator of ruin probability in a Lévy risk model
- The aggregate claims process $X = \{X_t, \geq 0\}$ is modeled by a pure-jump Lévy process
- Assume that high-frequency observed data on X is available
- The estimator is constructed based on Pollaczeck-Khinchine formula and Fourier transform
- Risk bounds as well as a data-driven model selection methodology are presented

Some related references

- Croux, K., Vervaerbeke, N. (1990). Nonparametric estimators for the probability of ruin. *Insurance: Mathematics and Economics* **9**, 127-130.
- Frees, E. W. (1986). Nonparametric estimation of the probability of ruin. *ASTIN Bulletin* **16**, 81-90.
- Hipp, C. (1989). Estimators and bootstrap confidence intervals for ruin probabilities. *ASTIN Bulletin* **19**, 57-70.
- Politis, K. (2003). Semiparametric estimation for non-ruin Probabilities. *Scandinavian Actuarial Journal* **2003(1)**, 75-96.
- Yasutaka S. (2012). Non-parametric estimation of the Gerber-Shiu function for the Wiener-Poisson risk model. *Scandinavian Actuarial Journal* **2012 (1)** 56-69.

Our model of surplus

$$U_t = u + ct - X_t,$$

where $u \geq 0$ is the initial surplus, $c > 0$ is the constant premium rate. Here the aggregate claims process $X = \{X_t, t \geq 0\}$ with $X_0 = 0$ is a pure-jump Lévy process with characteristic function

$$\phi_{X_t}(\omega) = \mathbb{E}[exp(i\omega X_t)] = exp\left(t \int_{(0,\infty)} (e^{i\omega x} - 1)\nu(dx)\right),$$

where ν is the Lévy measure on $(0, \infty)$ satisfying the condition $\mu_1 := \int_{(0,\infty)} x\nu(dx) < \infty$.

Ruin probability

The ruin probability is defined by

$$\psi(u) = \mathbb{P}\left(\inf_{0 \le t < \infty} U_t < 0 | U_0 = u\right).$$

Assumption *The safety loading condition holds, i.e.* $c > \mu_1$.

Some notation

Let L_1 and L_2 denote the class of functions that are absolute integrable and square integrable, respectively. For a L_1 integrable function g we denote its Fourier transform by

$$\phi_g(\omega) = \int e^{i\omega x} g(x) dx.$$

For a random variable Y we denote its characteristic function by $\phi_Y(\omega)$. Note that under some mild integrable conditions Fourier inversion transform gives

$$g(x) = \frac{1}{2\pi} \int e^{-i\omega x} \phi_g(\omega) d\omega.$$

Pollaczeck-Khinchine formula

Let

$$H(x) = \frac{1}{\mu_1} \int_0^x \nu(y, \infty) dy$$

with density $h(x) = \nu(x, \infty)/\mu_1$. Then the Pollaczeck-Khinchine type formula for ruin probability is given by

$$\begin{aligned}
\psi(u) &= 1 - (1-\rho) \sum_{j=0}^{\infty} \rho^j H^{j*}(u) \\
&= \rho - (1-\rho) \sum_{j=1}^{\infty} \rho^j \int_0^u h^{j*}(y) dy \\
&= \rho - (1-\rho) \int_0^u \chi(x) dx,
\end{aligned}$$

where $\rho = \mu_1/c$, $\chi(x) = \sum_{j=1}^{\infty} \rho^j h^{j*}(x)$.

The convolutions are defined as

$$H^{j*}(x) = \int_0^x H^{(j-1)*}(x-y)H(dy), \;\; h^{j*}(x) = \int_0^x h^{(j-1)*}(x-y)h(y)dy$$

with $H^{1*}(x) = H(x)$ and $h^{1*}(x) = h(x)$.

Estimate the parameter ρ

Suppose that the process X can be observed at a sequence of discrete time points $\{k\Delta,\ k = 1, 2, \ldots\}$ with $\Delta > 0$ being the sampling interval. Let

$$Z_k = Z_k^{\Delta} = X_{k\Delta} - X_{(k-1)\Delta}, \quad k = 1, 2, \ldots, n.$$

Assumed that the sampling interval $\Delta = \Delta_n$ tends to zero as n tends to infinity.

An unbiased estimator for ρ is given by

$$\hat{\rho} = \frac{1}{cn\Delta} \sum_{k=1}^{n} Z_k. \tag{1}$$

Estimate the function $\chi(x)$

By integration by parts

$$\phi_h(\omega) = \frac{1}{\mu_1} A(\omega),$$

where

$$A(\omega) = \int_0^\infty e^{i\omega x} \nu(x, \infty) dx = \int_0^\infty \frac{e^{i\omega x} - 1}{i\omega} \nu(dx)$$

is the Fourier transform of $\nu(x, \infty)$.

Standard property of Fourier transform implies that

$$\phi_{\chi}(\omega) = \int e^{i\omega x} \sum_{j=1}^{\infty} \rho^j h^{j*}(x)dx = \sum_{j=1}^{\infty} \rho^j \left(\phi_h(\omega)\right)^j = \frac{A(\omega)}{c - A(\omega)}.$$

Thus, Fourier inversion transform gives the following alternative representation for $\chi(x)$,

$$\chi(x) = \frac{1}{2\pi} \int e^{-i\omega x} \frac{A(\omega)}{c - A(\omega)} d\omega. \tag{2}$$

Remark

The denominator $c - A(\omega)$ is bounded away from zero because by $|e^{i\omega x} - 1| \leq |\omega x|$ we have

$$|c - A(\omega)| \geq c - \int_0^\infty \left| \frac{e^{i\omega x} - 1}{i\omega} \right| \nu(dx) \geq c - \mu_1 > 0$$

In order to obtain an estimator for $\chi(x)$ we can first estimate $A(\omega)$. Note that $\{Z_k\}$ are i.i.d. with common characteristic function

$$\phi_Z(\omega) = exp\left(\Delta \int_0^{\infty} \left(e^{i\omega x} - 1\right) \nu(dx)\right).$$

By inverting the above characteristic function we obtain

$$\int_0^{\infty} (e^{i\omega x} - 1)\nu(dx) = \frac{1}{\Delta} Log\left(\phi_Z(\omega)\right),$$

where *Log* denotes the *distinguished logarithm*

Using the fact that $A(\omega) = \frac{1}{\Delta}\frac{Log(\phi_Z(\omega))}{i\omega}$, we know that a plausible estimator is

$$\frac{1}{\Delta}\frac{Log(\hat{\phi}_Z(\omega))}{i\omega},$$

where $\hat{\phi}_Z(\omega) = \frac{1}{n}\sum_{k=1}^{n} e^{i\omega Z_k}$ is the empirical characteristic function.

However, on the one hand, the *distinguished logarithm* in the above formula is not well defined unless $\hat{\phi}_Z(\omega)$ never vanishes; on the other hand, it is not preferable to deal with logarithm for numerical calculation.

In order to overcome this drawback, we follow a different approach.

Write $A(\omega)$ in the following form,

$$A(\omega) = \frac{\phi_Z(\omega) - 1}{i\omega\Delta} + \frac{1}{i\omega\Delta}\left[Log(\phi_Z(\omega)) - (\phi_Z(\omega) - 1)\right].$$

Using the inequality $|e^{i\omega x} - 1| \leq |\omega x|$, we have $|\phi_Z(\omega) - 1| \leq |\omega|\Delta\mu_1$. Together with the inequality $|Log(1+z) - z| \leq |z|^2$ for $|z| < \frac{1}{2}$, we obtain

$$|Log(\phi_Z(\omega)) - (\phi_Z(\omega) - 1)| \leq (\omega\Delta\mu_1)^2, \tag{3}$$

provided that $\Delta|\omega|$ is small enough.

If $\Delta|\omega| \to 0$, $[Log(\phi_Z(\omega)) - (\phi_Z(\omega) - 1)]/(i\omega\Delta)$ can be neglected, i.e.

$$A(\omega) \approx \frac{\phi_Z(\omega) - 1}{i\omega\Delta}.$$

Hence, we propose the following estimator for $A(\omega)$,

$$\hat{A}(\omega) = \frac{\hat{\phi}_Z(\omega) - 1}{i\omega\Delta}, \tag{4}$$

where for $\omega = 0$ (4) is interpreted as the limit $\hat{A}(0) := \frac{1}{n\Delta}\sum_{k=1}^{n} Z_k$.

Write $E_n(\omega) = \{|c - \hat{A}(\omega)| \geq (n\Delta)^{-\frac{1}{2}}\}$. Replacing $A(\omega)$ in (2.3) by $\hat{A}(\omega)$ gives the following estimator

$$\hat{\chi}(x) = \frac{1}{2\pi} \int e^{-i\omega x} \frac{\hat{A}(\omega)}{c - \hat{A}(\omega)} \mathbf{1}_{E_n(\omega)} d\omega, \qquad (5)$$

where the indicator function $\mathbf{1}_{E_n(\omega)}$ is used to guarantee that the denominator is bounded away from zero.

A cut-off modification of estimator of χ:

$$\hat{\chi}_m(x) = \frac{1}{2\pi} \int_{-m\pi}^{m\pi} e^{-i\omega x} \frac{\hat{A}(\omega)}{c - \hat{A}(\omega)} \mathbf{1}_{E_n(\omega)} d\omega, \qquad (6)$$

where m is a positive cur-off parameter.

Estimator for ruin probability

$$\begin{aligned}\hat{\psi}_m(u) &= \hat{\rho} - (1-\hat{\rho})\int_0^u \hat{\chi}_m(x)dx \qquad (7)\\ &= \hat{\rho} - \frac{1-\hat{\rho}}{2\pi}\int_{-m\pi}^{m\pi} \frac{1-e^{-i\omega u}}{i\omega}\frac{\hat{A}(\omega)}{c-\hat{A}(\omega)}\mathbf{1}_{E_n(\omega)}d\omega,\end{aligned}$$

where the second step follows from Fubini's theorem.

Risk bounds

For $v \in L_1 \cap L_2$ let

$$\| v \|^2 = \int |v(x)|^2 dx.$$

Under some assumptions and assume that $\Delta \to 0$, $m\Delta \to 0$ and $n\Delta \to \infty$, then

$$\mathbb{E}\|\hat{\chi}_m - \chi\|^2 = O\left(m(n\Delta)^{-1} + m^{-2a}\right).$$

In particular, when $m = O((n\Delta)^{\frac{1}{2a+1}})$ and $n\Delta^{2a+2} \to 0$, we have

$$\mathbb{E}\|\hat{\chi}_m - \chi\|^2 = O\left((n\Delta)^{-\frac{2a}{2a+1}}\right).$$

Model selection

We know that the estimator depends heavily on the cut-off parameter m. We propose a data-driven strategy to choose m. We select adaptively the parameter m as follows:

$$\hat{m}^* = \arg\min_{m\in\{1,2,\ldots,m_n\}} \{\gamma_n(\hat{\chi}_m) + pen^*(m)\}. \qquad (8)$$

where

$$\gamma_n(v) = ||v - \hat{\chi}_m||^2 - ||\hat{\chi}_m||^2,$$

and

$$pen^*(m) = 96c^2 \frac{1/n\Delta \sum_{j=1}^{n} Z_j^2}{(c - 1/n\Delta \sum_{j=1}^{n} Z_j)^4} \frac{m}{n\Delta},$$

if $|c - 1/n\Delta \sum_{j=1}^{n} Z_j| \geq \epsilon_n$, and

$$pen^*(m) = \frac{m}{n\Delta},$$

if $|c - 1/n\Delta \sum_{j=1}^{n} Z_j| < \epsilon_n$.

Simulation studies

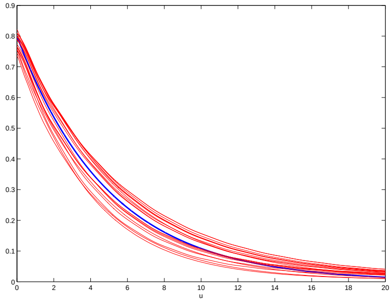

Figure: Estimation of the ruin probability in the compound Poisson risk model with exponential claim sizes. True ruin probability (blue line) and 20 estimated curves (red lines). Sample size $n = 1000$, sampling interval $\Delta = 0.01$.